

Documenting Variation in (Endangered) Heritage Languages: How and Why?

Naomi Nagy
University of Toronto

This paper contributes to recently expanded interest in documenting variable as well as categorical patterns of endangered languages. It describes approaches, tools and curricular developments that have benefitted from involving students who are heritage language community members, key to expanding variationist focus to a wider range of languages. I describe aspects of the Heritage Language Variation and Change Project in Toronto, contrasting a “truly” endangered language to a less clearly endangered language. Faetar, with <700 homeland speakers (in Italy) and some 200 in Toronto, and no transmission to a third generation in Toronto, is endangered by any definition. Heritage Italian, in contrast, is a diasporic variety related to a robust homeland variety as well as the mother tongue of 166,000 Torontonians. However, reports of strong English influence on the language and transmission statistics both suggest that it too is endangered in Toronto. Homeland and Heritage patterns are compared to better understand the processes of language variation and change in lesser-studied varieties, with a focus on null subject patterns. Analysis of the more endangered language helps interpret otherwise ambiguous patterns in the less endangered language. Results indicate that neither heritage language exhibits the simplification anticipated for small languages in contact with a majority language.

1. INTRODUCTION. Recently, there is expanded interest in the documentation of variable as well as categorical patterns in the process of endangered language documentation, as encapsulated in this recommendation by Childs et al. (2014:1):¹

[The field would benefit from] a greater emphasis on conversational data and the documentation of naturally occurring conversation ... collecting sociolinguistic data which can inform language planning and policy; and creating opportunities for training in sociolinguistic documentation.

This quotation succinctly describes the methods of variationist (or Labovian, cf. Labov 1972, 1984) sociolinguistics, an approach that seeks to understand the factors that

¹ I am indebted to the many speakers who contributed time, expertise, and data to this project; the research assistants who collected, curated and analyzed the data (listed at http://projects.chass.utoronto.ca/ngn/HLVC/3_2_active_RA.php and http://projects.chass.utoronto.ca/ngn/HLVC/3_3_former_RA.php); SSHRC grants 410-2009-2330, 435-2016-1430; the editors of this issue for encouragement to prepare this article; and Nathalie Dion, Katie Drager, Rick Grimm, Shana Poplack, and Luke Zhou helpful input.

constrain variable patterns (phonological, syntactic, etc.) in conversational speech. Importantly, we also are observing expanded work in the opposite direction: variationist sociolinguists are increasingly examining smaller and less well-documented languages, as evidenced by the LSA Special Session from which this issue arose (cf. Drager et al. 2016, Chirkova et al. 2016) and works in Harrison (2005), Meyerhoff & Nagy (2008) and Stanford & Preston (2007). This work is critical for understanding whether the types of patterns that have been reported thus far in variationist sociolinguistics apply to a broader range of languages and communities. Such an approach can highlight patterns of individual and dialect differences rather than forcing Nagys of grammars to choose between variants (see Nagy 2009). Such documentation can be valuable to language internal comparative reconstruction, adding value to early phases of language documentation.

This paper has two intertwined goals. The first is to show a useful adaptation of sociolinguistic methodology that makes it possible to study under-documented varieties in a variationist framework: training speakers of lesser-studied languages as sociolinguists. The second is to illustrate this process, and its benefits, by describing the methods and outcomes of a study comparing patterns of null subject (pro-drop) variation in Homeland and Heritage varieties of two languages, Faetar and Italian. In addition to documenting these varieties, this contributes to better understanding the processes of language variation and change in lesser-studied varieties, in order to see how well generalizations based on larger and better-documented languages can account for new types of data. The cross-variety comparison helps interpret otherwise ambiguous patterns, suggesting an additional benefit of expanding our focus to additional languages: the potential to use patterns and findings from one language to help understand another.

Adapting methods from one field to another should not be undertaken without considerations of appropriateness. While the Childs et al. quote above seems to recommend sociolinguistic methods, their immediate continuation questions whether the western origins of the field may be problematic in efforts to expand to other (e.g., non-western, non-first world, non-majority) languages:

Consideration of sociolinguistic language documentation [...] questions the ways in which Western language ideologies [...] may be unduly influencing documentary practice in other parts of the world (Childs et al. 2014:1).

This highlights a critical question that arises when we apply comparative sociolinguistic methods to a bigger range of languages: What differences matter? That is, what exactly changes when we look at minority and smaller languages, languages that are less well-documented, both generally and in terms of the variation they encompass? We see this addressed in Labov (2015), which describes how explorations in languages beyond English reinforce but also better focus our understanding of basic sociolinguistic principles, particularly related to the role of gender. However, much of our understanding of the connections between synchronic variation and change remain based on work from a few large, well-studied languages.

This paper straddles the domains of languages less studied in the variationist paradigm and of endangered language documentation, with the aim of modeling a

combined approach. In the remainder of this introduction, I summarize the most relevant distinctions in methods (of data collection and the inquiry processes) between variationist sociolinguistics and endangered language documentation.

2. AN OVERVIEW OF SOCIOLINGUISTIC INQUIRY. Nearly fifty years of work in the Labovian paradigm of simultaneous quantitative analysis of multiple factors influencing variation in spontaneous speech (cf. Labov 1984 for detailed description of methods) has led us to expect certain social patterns to repeatedly emerge. These involve the correlation of social factors like age, ethnicity, sex and status² to linguistic variation. We thus understand the variable linguistic patterns correlated to these factors to be indicative of the social structure of the speech community: speakers with more social factors in common will share greater similarity in how they talk. We also have developed an understanding of patterns of social differences, reflected in linguistic patterns, that are predictive of variables undergoing change vs. those variables which are stable over time and that distinguish between changes from below and changes from above (cf. Labov 2001).³ Some examples of generally accepted principles or generalizations, mostly excerpted from articles in a recent handbook (Bayley et al. 2013) are listed in (1). To exemplify the approach, age and sex-related patterns are considered in the analysis of pro-drop variation that follows.

(1) *Widely-accepted sociolinguistic generalizations regarding social factors*

- Age-related patterns
 - Apparent Time Construct (Bailey et al. 1991): Younger speakers represent newer stages of a variety than older speakers
 - Rate of use of innovative variants peaks in the adolescent age range (D'Arcy 2013:490)
 - Vernacular reorganization by young adults (Labov 2001)
 - Stability across the (adult) lifespan (Walker & Meyerhoff 2013:178, Sankoff 2013:261, but see D'Arcy 2013:489 and Sankoff 2013:270 for counter-examples)
- Sex-related patterns
 - Women use fewer stigmatized variants than men (Bayley 2013:14)
 - Women are the innovators in linguistic change (Labov 1990:205)
 - Women pattern like men of the next-higher social class (Bayley 2013:14)
- Class-related patterns
 - Curvilinear social class pattern: we find more innovation among speakers of central than peripheral social classes (Labov 1990:205)
 - Hypercorrection by (females in) the second-highest social class (Labov 2001:223-224)

² Status is often operationalized by measures of socio-economic standing, education, or type of employment.

³ Changes from below are those that arise without conscious awareness of the speakers producing them and often are exhibited first by speakers with lower status. In contrast, changes from above are produced with greater awareness and are used first by speakers with higher status.

- Style-related patterns
 - Convergence of rates of use of stigmatized vs. non-stigmatized forms in more formal styles (Labov 1972)
 - Bell's Axiom: stylistic stratification reflects social stratification in a particular community for a particular variable (Bell 1984)
- Network-related patterns
 - "relatively closed social networks with strong community-based ties tend to maintain non-legitimized linguistic codes... [while networks of people with] "weak community-based ties tend to accept the dominance of legitimized linguistic codes" (Milroy & Milroy 1992:22)
 - Shared repertoires within *communities of practice* (Holmes & Meyerhoff 1999)

We have also developed expectations regarding linguistic factors, as discussed in Guy (2015). We expect, for example, phonological variation to be gradiently influenced by phonotactic constraints, markedness, and satisfaction of the Obligatory Contour Principle. We expect syntactic variation to be influenced by information structure and ambiguity effects. In the analysis below, information structure is operationalized and ambiguity is indirectly examined through the coding of person/number and form of the subject, and tense, of each token sentence.

These patterns of effects of social and linguistic factors are critical components of the variationist field, often received as universals. But, unlike the field of linguistics more broadly, variationist sociolinguistic research has developed by examining primarily English, with research on French and Spanish occupying most of the rest of the publication space of this field (Nagy & Meyerhoff 2008). Across even these three widely-studied languages, there is not a great deal of overlap in what sort of variation is investigated. That is, different linguistic patterns have become the "showcase" variables for different languages. For example, variation in pre-verbal negation marking, future tense marking, affrication of coronal stops and deletion of word-initial /l/ have been extensively investigated in French. None of these have been investigated (much) in English, which instead has been the target of much study of variation in the copula, verbs of quotation, velar fronting (ING),⁴ and consonant-cluster simplification (T,D). Spanish has much work devoted to null-subject variation and aspiration/deletion of word-final /s/. While, of course, other variables have also been examined, listing them here would not increase the overlap in targets of investigation across these languages.

In addition to a difference in dependent variables examined, there is not consensus on the effects of the independent variables. As we look toward smaller languages, language contact issues cannot be ignored. In a world where most individuals speak more than one language and most languages are in contact with other languages, the study of language in its social context must take into account language contact effects

⁴ Abbreviations used in this paper: HLVC Heritage Language Variation and Change Project; (ING) sociolinguistic variable consisting of the alternation between velar and alveolar nasals in the *-ing* morpheme; ISTAT Istituto Nazionale di Statistica; LSA Linguistic Society of America; NWAV New Ways of Analyzing Variation conference; (T,D) sociolinguistic variable consisting of the variable deletion of a word-final /t/ or /d/ in a consonant cluster; VOT Voice Onset Time.

(Ravindranath 2015:243). To understand whether constraints on variation, such as those listed in (1), reflect cross-linguistic universals, it is thus necessary to look more closely at a wider range of languages and linguistic contexts, including non-western, smaller and lower-status languages.

There are, however, a number of reasons that such work is not happening more quickly. What we might consider as “classical” sociolinguistic training involves work on large, well-documented languages and differs in crucial ways from documentation work. Indeed, in many cases, student-training involves contributing to an already-existing project and corpus of data – the language is not just well-documented in the classical sense, but also in terms of its variation. Furthermore, training tends to involve languages whose grammar is well-described, often including even the variation of the “practice” variables that students examine – think of the many student projects on (ING) or (T,D). Fieldworkers and analysts are often native speakers of the language under investigation. There is usually a big and accessible population of speakers of the language. In addition to the language and many of its variable patterns being already well-described, such training is often undertaken in communities whose social structures are well described. A final critical fact is that the speakers are often treated as monolingual – descriptions, particularly in studies of North American English, often don’t even mention the possibility that they may also speak another language.

2.1 Working with smaller languages. Scholars whose goal it is to document smaller languages face an entirely different scenario. I describe first some of the differences between working with dominant vs. minority languages, and then focus in on additional differences that arise in work with endangered languages. An underlying presumption of much research about perceptions of (and perhaps also production of) linguistic variation is encapsulated in Herder’s equation, the concept of a necessary connection between one’s language, one’s culture, and one’s nation-state. For speakers of minority languages, such connections may not exist (Foley 2005:157-8). It is important for us to learn how this may influence attitudinal factors.

First, it is important to recognize speakers of minority languages are frequently bilingual and that bilingual⁵ speakers have their very choice of language as a resource for constructing identity, as explained by Wei (2013:43).

For the multilingual speaker, language choice is not only an effective means of communication but also an act of identity (Le Page & Tabouret-Keller 1985). Every time we say something in one language when we might just as easily have said it in another, we are reconnecting with people, situations, and power configurations from our history of past interactions and imprinting on that history our attitudes towards the people and languages concerned. Through language choice, we maintain and change ethnic-group boundaries and personal relationships, and construct and define ‘self’ and ‘other’ within a broader political economy and historical context.

⁵ ‘Bilingual’ denotes bi- or multilingual.

Therefore, bilinguals need not rely solely on language-internal variation (of the types listed in (1)) for identity-construction work (see discussion in Nagy et al. 2016). The impact of this could best be understood by studies that look both at speakers' choices between languages and within each language – not something often seen in the literature – and then make connections between these. Results so far, however, suggest that there is little correlation between these types of choices. Poplack & Turpin (1999:158) report, “no correlation between variant use [future tense forms in Ottawa-Hull French] and level of bilingualism was found.” Working with the same corpus, Leroux & Jarmasz (2006:11-12) and Lealess & Smith (2011:12) report no effect of the Cumulative French Proficiency Index on null subjects and subject relative pronoun choice, respectively. This index includes frequency of use of French and English in its calculations (Poplack et al. 2006:191). Grimm (2015:141-2) shows a similar lack of correlation between production of subjunctive forms and relative use of French and English among bilinguals in four Ontario towns. Similarly, Nagy et al. (2014) explored ways of showing a connection between (self-reported) language use and several sociolinguistic variables in the heritage language production of various groups of bilingual speakers and was not able to show significant correlation (this includes the Italian null subject variation examined in §4.1). In contrast, however, Waltermire (2010:296) reports a significant effect of rate of use of Portuguese *vs.* Spanish on the pattern of production of a Portuguese-source variant in bilingual speakers' Spanish and Nagy et al. (2014) did find a correlation between (self-reported) language use and null subject variation for first generation Heritage Cantonese speakers).

The following distinctions between variable patterns in minority *vs.* dominant languages may (or may not) be related to this role of language choice in identity construction and reflect the non-universality, when we expand our purview to include minority languages, of the claims listed in (1). Stanford & Preston (2007:8) note that social factors may behave differently in minority languages, noting counterexamples to the trend that women lead linguistic changes. Blainey (2013) reports that social variables “like sex, socio-economic class, literacy, ethnic identity, speech register, and home town” may lose influence in contracting languages, citing work in Dubois & Noetzel (2005:135), King (1989), and Rottet (1995:309). She further suggests that education (not necessarily in the language being studied) may mediate the role of other social factors in language minority situations. Age, also, may operate differently. For example, studies of English have suggested an ‘adolescent peak,’ the phenomenon that late teens show higher rates of use of innovative variants than speakers both older and younger than them (Cedergren 1973, 1988, D’Arcy 2013, Labov 2001). In the western urban societies where this has been reported (cf. Tagliamonte & D’Arcy 2007), this age effect may relate to moving out of the parental home and mixing more with people from other communities in the university setting.⁶ In communities where leaving home for university in late adolescence is not a dominant trend, or where the language under study is not the language of education, we do not know if such a linguistic pattern would be observed. As one potential counterexample, youth around age 12, living in Faeto, Italy, begin to speak their hometown variety, called Faetar (discussed in Section 4), more frequently than

⁶ In a recent survey, 75% of Canadians had some access to post-secondary education by the age of 21 (McMullen 2011).

younger children (Nagy 2000:6). This occurs at the time when they must commute to a neighboring town for school, where they mix with students who are not speakers of Faetar. The effect of this lifestyle on variable patterns in their speech has not yet been examined. Heritage languages are another ripe context for study of this effect: the set of interlocutors with whom speakers interact in their heritage language likely changes less, with the start of university studies, than the set of interlocutors with whom English (or the majority language) is used. This is but a few examples of how the “universals” in (1) might be challenged by studies of additional (types of) languages.

Additionally, to the extent that documentation is not conducted by speakers of the target language, communities are likely to have different social structures than those with which the researchers are familiar from personal experience. The challenges of calibrating linguistic variation to social variation in this situation are evident (cf. Stanford 2013). A framework for cross-cultural comparison is lacking in our field.

Next, attitudes toward a minority language that is not (well-) supported by state institutions, etc., have a greater likelihood of being negative than that of a well-supported majority language. As Garrett (2010:7) puts it:

Attitudes toward language varieties are generally influenced by the process of language... When one language variety is deemed “standard,” that impacts (usually devalues) the status of the other language varieties and of the social groups using those language varieties in a particular context.

Milroy (2001:535) suggests this causal relationship:

Prestige is acquired through the process of standardization in a given society in a given time, and it is most often associated with a social group with weighty influence in the society.

Ó Riagáin (2008:333-4) and Hilton & Gooskens (2013:150-153) provide additional evidence of this trend. This regards both the attitudes of the minority language’s speakers and those with whom they may be in contact. As we know that speaker attitudes are correlated to at least some aspects of linguistic variation, this may impact linguistic variability. To illustrate, here is a typical description of heritage Italian. While these quotations are from a linguist, they represent the view expressed by many speakers in my own fieldwork experience in several languages and communities. Two common themes are expressed: that heritage or diaspora varieties have eroded due to simplification of the grammar and that such varieties are necessarily subject to interference from the dominant language in the community.

È evidente la semplificazione non solo del sistema verbale, ma più in generale la semplificazione delle strutture linguistiche, ... la perdita di alcuni tratti (soprattutto a livello morfologico) (Palumbo 2014:31).
Simplification is evident in the verbal system and more generally in linguistic structure, losing certain features, especially morphological.

L'italiano di emigrazione si presenta come ... una varietà molto simile all'italiano popolare, ...soggetta all'influenza della lingua di adstrato (Palumbo 2014:23).

The Italian of immigration is very similar to *Italiano Popolare*, subject to adstrate influence.

Finally, in language contact situations, speakers may differ in their degree of fluency in the language(s) being documented, requiring the researcher to make judgments about the representativity of the data that different speakers provide (cf. Nagy 2009:411). Speakers of heritage/diaspora varieties may not fully replicate homeland varieties because of “quantitatively deficient input” to learners. That is, heritage speakers receive less input in their heritage language than those living in a monolingual community to the extent that they receive input in the community’s dominant language. While much work has suggested that this leads to different grammars (cf. Montrul 2009, Polinsky 2011), it is not clearly established whether this also influences sociolinguistic competence (but see Albirini & Chakrani 2016, Kang 2015) for recent work in this area. Because of the possibility of language change accelerated due to contact phenomena, these can be particularly fruitful sites of investigation to improve our understanding of linguistic variation.

2.2 Working with endangered languages. We turn next to the additional challenges faced by researchers documenting variation in languages which are not just smaller but also endangered. First, of course, the targets of language documentation are, by definition, languages that are not well-documented. Basic facts about the grammar must be established, often simultaneously with assessing the variable patterns (Meyerhoff *forthcoming*). “Pre-fab” variables relevant to these languages have not been previously described in publication, so they are not available for diagnostic purposes. For moribund languages, it may not be possible to sample the language of every speaker category (by, e.g., age, sex or status) in order to categorize variable patterns as likely to be indicative of stability, change from above or change from below, as discussed in (1).

Second, there are frequent claims of conservatism in heritage and/or endangered languages (cf. Woolard 1989:355). We can find these in the literature at least since Dorian’s (1980) pioneering work with semi-speakers of Eastern Scots Gaelic. Conservatism was explored in Thepboriruk (2015) and was the focus of the “Heritage and diaspora varieties” panel at the 2015 NWAV44 conference.

Third, with fewer speakers available, the importance of careful consideration of inter-speaker variability increases. In addition, when there is no “standard” to calibrate against, documentation is subject to the impact of inter-transcriber, inter-translator and inter-coder variability, in ways that require further exploration (Rice 2014; Julien Carrier, p.c., Dec. 21, 2015). Thus recording a range of speakers and applying variationist methods to account for as much of the variation as possible increases in importance.

This section has highlighted some of the distinctions between the research contexts that are typical of variationist work and those typical of endangered language documentation. After considering these, we must consider whether, differences

notwithstanding, we could accept the “universal principles” in (1) as guiding principles in any case.

3. THE HERITAGE LANGUAGE VARIATION AND CHANGE PROJECT. A key to expanding the set of the languages that can be documented and analyzed in this way is broader student involvement. In many universities today, particularly those catering to urban populations in multilingual cities, the pool of languages spoken by students is considerably larger than the pool spoken by sociolinguistics professors. In such contexts, our field can benefit by encouraging and training students to conduct research on more languages, expanding the range of languages subjected to variationist analysis and comparison. This section highlights and defines the critical role of students and suggests means of increasing student involvement in minority language documentation.

The Heritage Language Variation and Change project (HLVC, Nagy 2011) is a large-scale project investigating variation and change in heritage languages spoken in the Greater Toronto Area (hereafter, Toronto). In order to push variationist research beyond its majority language and monolingual speaker-focused core, its goals are to document and describe a set of heritage languages. Languages currently included are Cantonese, Faetar, Hungarian, Italian, Korean, Polish, Russian and Ukrainian. Of these, only half are spoken by faculty in the Nagy’s Linguistics Department. This expansion of variationist approaches provides evidence that bears on the issues raised in the introduction: whether the “universal principles” in (1) are applicable in minority and lesser-studied language contexts.

3.1 Overview of HLVC Methods. Documentation involves the creation of a corpus of spoken language, transcribed and time-aligned to audio recordings, that is available online for research on a variety of topics. Our in-depth interviews with HL speakers simultaneously produce speech samples containing tokens of many linguistic variables and emic information about potentially relevant social contexts.

The documentation process entails comparison of variable patterns in each heritage variety, their homeland comparator languages, and the local variety of English. This allows us to compare both stable variation and changes in progress found in these varieties to those reported in the literature for dominant and better-described languages. Thomason & Kaufman (1988:111) highlight a vexing paradox in language contact research: once contact has occurred, it may not be easy to access the pre-contact variety, yet contrasting these is crucial. To trace the trajectory of change, we contrast the linguistic systems of speakers with greater and lesser contact with their HL and the majority language (English in Toronto), providing snapshots of language change on a “fast track.” These are interpreted and compared in the variationist paradigm to determine whether there is contact-induced variation (Poplack & Levey 2010). In addition to comparing heritage speakers who report greater vs. lesser use of English, we compare heritage speakers to homeland speakers. This latter group, even in today’s globalizing world, have less contact with English than do people living in Toronto. This provides a clear backdrop against which contact-induced changes, if any exist, should be highlighted (cf. Thomason 2001, Nichols 2008). As is typical in variationist investigation, patterns of variation related to age and sex, as summarized in (1), are also sought.

3.2 Student involvement in the HLVC Project. Additional goals of the HLVC Project include training students and speakers of the heritage languages to conduct linguistic research and providing resources to speakers and communities (cf. Nagy 2017 for additional details). Working directly with students who are community members allows for a more ethnographically-informed approach, partially alleviating the problem described above of social structures that differ from those in the researcher's personal experience. Students who are community members provide a conduit to communities that might otherwise be inaccessible to outsiders. As noted above, students are fluent speakers of a wider range of languages than those spoken by sociolinguistics professors (within a department), a key asset for data-collection in the variationist sociolinguistic paradigm.

The HLVC Project has fine-tuned methods allowing a large network of students and faculty with complementary expertise to efficiently collect, transcribe, and archive data. Our heavy reliance on student involvement is motivated by two goals. The first is to provide training to students (and allow us to benefit from university resources designated for this purpose). The second is to promote research on languages not spoken by professional sociolinguists in Toronto. Research methods, tools and principles, as well as courses and programs of independent study at various levels, have benefitted from the involvement of students who are members of each heritage language community. Students are regularly consulted regarding methodology, allowing the project to benefit from technological innovations that students may introduce as well as utilizing methods that student researchers find accessible and beneficial. This has led to substantive innovations in our practices and new research questions.

This project is located in Toronto (pop. ~5.5 million). Half its current residents were born outside of Canada and thus many languages are spoken. The 2011 census reports 200+ home languages. (Statistics Canada 2011a). The University of Toronto has nearly 85,000 students. One third of undergraduate students (in a recent University sample) report that they speak a language other than English at home.⁷ The University encourages undergraduate research, a critical advantage to the HLVC project. Thus there are university resources to support student involvement in research projects. Recruitment of such students comes in many guises. Most years, I offer a first-year seminar, "Exploring Heritage Languages," that is aimed at introducing research questions and methods to speakers of heritage languages (and other interested students). Workstudy positions and RAships are offered that require being a fluent speaker of a heritage language but no linguistic skills. Posters recruiting participants often also attract the attention of new researchers. Students are attracted primarily because of opportunities to engage with and support their heritage languages, but they also are compensated through paid work; volunteer citations in their co-curricular transcripts; or academic credit for research completed as part of a course, independent student, or the undergraduate Research Opportunity Program.

The many students who have participated as RAs are featured on the project's website.⁸ Students in every stage of their bachelor's degree studies, as well as MA and PhD students, participate. This range of methods of attracting students to the project has produced a trajectory where a few students have worked on the HLVC project throughout

⁷ <http://leadershipstudy.net>

⁸ http://projects.chass.utoronto.ca/ngn/HLVC/3_2_active_ra.php,
http://projects.chass.utoronto.ca/ngn/HLVC/3_3_former_ra.php

their entire undergraduate (or MA) experience, or for the last years of their BA plus their MA and, in one case, also the PhD.⁹ However, the majority of student-researchers (upwards of 100 since 2009) have been involved in only one course or research assistantship. Students who show particular interest or aptitude are encouraged to continue.

Via this array of opportunities, we train students with high HL proficiency to participate in every stage: designing methods, recruiting and interviewing speakers, transcribing and coding data, analyzing sociolinguistic patterns within and across languages, and dissemination. Students with training in sociolinguistics, or in the process of acquiring such training, but with less or no HL proficiency are also involved in the analysis and dissemination stages, working in collaboration with fluent HL speakers. This develops transferable analytic, organizational and communicative skills of benefit to students across disciplines. Different stages of the process require different skills, allowing the project to offer opportunities to virtually all interested students. We maintain a stream of fresh input to the project with over a dozen students/year engaged in the project, building a trajectory of professionalization for speakers of less frequently studied languages.

Students who belong to each heritage language's community recruit, interview and record speakers, transcribe and proofread, produce meta-data to make the corpus more widely useful and accessible, and code variable linguistic patterns. Each of these steps may be conducted by different individuals or teams of students, making it essential that we have clear documentation of each stage of our work. This makes the HLVC Project run more smoothly but also produces data that is more readily usable by others.

Here it is relevant to note the exceptional case of Faetar, the one language in the HLVC Project that is endangered in its homeland variety, in addition to lacking transmission to a third generation in Toronto, coupled with a lack of continued migration from Faeto. This is the only language in the HLVC Project where there is no overlap between university students and heritage language speakers. Therefore, fluent speakers who are not students work (in paid or more often volunteer capacities) in tandem with students who are non-speakers of Faetar. This provides training to some community members who are not university students (often retired teachers) and also provides training more typical of documentation fieldwork to students who are not speakers of the language. Non-student community members and non-community-member students with knowledge of related varieties have successfully transcribed, coded and helped analyze the Faetar data used in this paper.

We develop, again with student input, assignments and class activities related to heritage language study.¹⁰ These benefit from a purpose-built, student-oriented data server.¹¹ Quotes illustrating positive student responses to course assignments and brief descriptions of the assignments are included in Nagy (2016:26-28).

In addition to weighing in on methodological decisions, students propose variables to analyze. As we lack local linguists working on some of these languages, this

⁹ These methods are listed online at http://projects.chass.utoronto.ca/ngn/HLVC/5_1_opportunities.php

¹⁰ See assignment descriptions at <http://individual.utoronto.ca/ngn/teaching.htm> (courses LIN251 and TBB199).

¹¹ Server described at <http://corpora.chass.utoronto.ca>.

is a critical contribution that allows the project to move forward. Students are also the principal creators of heritage language resources, cf. the Wikipedia page “Heritage Languages in Toronto,” created in 2016 by a group of first year undergraduate students, outlining the ethnolinguistic vitality (based on demographics, institutional support and status, cf. Giles *et al.* 1977) of nine heritage languages. Students have also proposed several ways of adapting methods developed by Walt Wolfram and his colleagues to bring awareness of linguistic variation into the heritage language classroom.

4. A COMPARISON OF TWO VARIETIES – MORE AND LESS ENDANGERED.

This section presents examples of the type of linguistic analysis that is conducted in the HLVC project, highlighting the blend of documentation and variationist analysis, and contrasting the methods required for bigger and smaller, more and less described, languages. Analysis of the more endangered language provides a new lens through which to interpret the variable patterns of the less endangered language. Unlike many previous studies of null-subject variation, here there is no evidence of influence from contact with local majority languages. The role of students in each stage is highlighted.

The two languages are Italian (the variety spoken in Calabria) and Faetar, a Francoprovençal isolate spoken in Faeto and Celle, two mountaintop villages in Apulia, Italy.¹² Italian was selected for inclusion in the HLVC Project because it was, at the time, the most common mother tongue in Toronto after English and has been a solid presence in the city for over 100 years. In contrast, Faetar was selected because so little is documented about both the language and the community, in Toronto as well as in Italy. We can expect certain differences between it and Italian and between both languages and English.

Faeto had 638 residents in 2014 and Celle had 166 (City Population 2016). Heritage Faetar demographics have not been documented by official census. It is estimated that there are between 30 and 300 speakers in Toronto, though estimates go as high as 1,000 for the number of people who trace their families to Faeto or Celle and now reside in Toronto. Most Faetani and Cellesi came to Toronto, directly from Italy, in the 1950’s and 1960’s. I have found no report of speakers arriving earlier. All Faetar speakers in Italy also speak Italian and, as far as I know, this is also the case for all Faetar speakers in Toronto. In Faeto, the number of English speakers is very close to zero.

In contrast, the census records indicate that 166,415 people, or ~3% of Toronto, reported Italian as their mother tongue in the 2011 census and 475,090 (~10% of the city) claim Italian as their ethnic origin (Statistics Canada 2011a,b). The situations of Faetar and Italian contrast in Italy as well. There, nearly 61 million people (virtually every resident) speak Italian. 35% of Italians living in Calabria report speaking Calabrese (a regional variety) as well as Italian (ISTAT 2007). We do not have access to statistics on the number of Torontonians from these groups who also speak English.

Heritage Italian speakers are commonly assumed to exhibit interference from English in their grammar, as indicated in the quotes from Palumbo above. Such interference has been documented for the lexicon (Cieri 1985, Danesi 1983-84, 1984, del Torto 2008, 2010, Fortier 1991). Danesi (1984), in contrast, suggests no influence of

¹² Faeto is 10 km from Celle by car, but less than 2 km on foot. I refer to the varieties spoken in both villages as Faetar, although they are sometimes referred to by distinct labels (Faetar, Cellese).

English on Heritage Italian phonology, morphology and syntax. This claim is supported in an analysis of VOT (voice onset time) of Italian-speakers in Toronto (Nagy et al. 2013) where no inter-generational differences indicating change toward the longer lag VOT pattern of English was noted. Faetar speakers, in addition, are expected to exhibit interference from Italian, both in Italy and in Toronto.

4.1 Case study: Variable Null Subject Presence (Pro-drop). To determine whether these two minority languages undergo interference from English at the morphosyntactic level, analyses of variable null subject presence, in English, Faetar, and Italian were undertaken. Variable null subject presence constitutes a “conflict site” (Poplack & Meechan 1998) in that the probabilistic grammar describing the rate of use and factors conditioning its use differ across these languages. English has virtually no null subjects in conversational speech (~2% of finite clauses, reported in Nagy et al. 2011). Faetar and Italian both have a much higher rate of finite verbs with no overt subject, though different factors condition the pattern in each language. Thus, if we find increasing similarity across generations among the languages which are in contact with each other, we have evidence that might be accounted for by the influence of language contact. In the contexts we consider, the overt and null forms appear to be in free variation, though stochastically conditioned by a number of factors. The variability is evident in (2) and (3), sentences produced by the same Heritage Italian speaker:

(2) **No overt subject pronoun (Italian)**

Ø Avevo 14 anni e mia moglie ce ne aveva 13.

Ø (I) was 14 and my wife was only 13... [I1M75A]¹³

(3) **Overt subject pronoun (Italian)**

Io ho detto “Ok, ce la faccio, io.”

I said, “OK, I’ll do it.” [I1M75A]

Faetar is more complicated in that sentences may contain zero, one or two (a weak plus a strong, personal or impersonal) subject pronouns. For this analysis null pronominal subjects are contrasted to any combination of one or more overt personal or impersonal pronouns. We thus have sentences of the types shown in (4-8), produced by Heritage Faetar speakers (adapted from Nagy et al. *forthcoming*). In addition, weak and strong subject pronouns may appear in a clause that has a [+Argument], or nominal, subject, as in (9) or with a demonstrative subject pronoun as in (10).

(4) **No overt subject pronoun (Faetar)**

/ε lu dʒore Ø stav a la kaz/

and that day, [Ø=I] was at the house (F1F79)

¹³ Speakers in the HLVC corpus are identified by speaker codes which indicate, in order, the first letter of the language, their generation, their sex and their age. The final character differentiates between otherwise demographically identical participants. Thus this first example comes from a Heritage Italian, first generation, male speaker who is 75 years old.

- (5) **Weak pronoun (Faetar)**
 /e i stavə vakant/
and it was vacant (F1M92A)
- (6) **Strong pronoun (Faetar)**
 /no iΛe sta tut:ə/
No, he was always... (F1F79A)
- (7) **Strong + Weak pronoun (Faetar)**
 /iΛe i e lu me prɛfɛri:tə/
She-strong she-weak is my favorite (F1F79A)
- (8) **Expletive pronoun (Faetar)**
 /o sundə kuntə k e pa lua:/
There are tales that are not true (F0M30A)
- (9) **Noun + weak pronoun (Faetar)**
 /pəke as a la metá də dʒunjə lə tʃɛras i amaturavand/
because in the middle of June the cherries they-weak ripen (F1M76A)
- (10) **Demonstrative pronoun + weak pronoun (Faetar)**
 /səl i fati a lu koriərə kanades/
This-one he works at the Corriere Canadese (F2M58A)

The investigation described below is a cross-linguistic comparison of both rates of null subjects and contextual effects that condition the probability of a null subject. Homeland and heritage, Italian and Faetar are examined via parallel methodology. Because of potential differences in the influence of the level of formality or accommodation toward the variable null subject pattern of different interviewers on the variable null subject rate, comparison of conditioning effects is established as a more robust way of determining degree of similarity between varieties under comparison (Torres Cacoullos & Travis 2015:376). In the context of this paper, though, it should be noted that the expectation of formality or accommodation effects is based on studies of larger languages (Tagliamonte 2006, Walker 2010) and is, thus, within the set of accepted “universals” that should be verified in a wider range of languages.

4.1.2 Methods. Participants in the HLVC Project are selected in order to fill a multi-dimensional matrix in which language, generation, age group and sex intersect, in order to seek effects of the type listed in (1). We define three generational groups of heritage speakers, as in (11):

- (11) Generations compared
- *Homeland:* The speaker was born and still resides in their hometown (in Calabria or in Faeto) in Italy.

- *First:* The speaker was born in Calabria or Faeto and moved to Toronto after age 18. They have subsequently lived in Toronto for 20+ years. Thus they fully acquired their mother tongue prior to immigration and have likely stabilized in the acquisition of additional languages.
- *Second:* The speaker was born in Toronto (or came from the homeland before the age of 6). Their parents qualify as 1st generation according to the definition above. They claim the heritage language as their mother tongue, having acquired it in the home from birth.
- *Third:* The speaker was born in Toronto. Their parents qualify as 2nd generation according to the definition above. Such speakers have been recorded for Italian but not Faetar.

For Italian, students who are members of the Heritage Italian speech community are in the process of recording speakers in each of the four age groups listed in Table 1, within each generation. Within each age by generation cross-section, we aim to record two males and two females. To provide comparable homeland Italian data, three students conducted interviews in Calabria, Italy as part of a study-abroad opportunity provided by their university. Fieldwork was preceded by several weeks of training in sociolinguistic goals and methods. The same sampling method is in place for Faetar, except that the Homeland Faetar speakers were recorded 1992-1994, prior to the conception of the HLVC Project. The speakers who could be analyzed in the current study are distributed as in Table 1. A known limitation of conducting fieldwork on smaller languages is it is harder (sometimes impossible) to collect speakers in some age-by-generation combinations (cf. Meyerhoff 2016:86). In this case, there are no Heritage Faetar speakers younger than 21 in Toronto. The Homeland distributions are considerably more complete.

	Age group	60+		40-59		21-39		< 21	
	Sex	M	F	M	F	M	F	M	F
Heritage Italian (n=11)									
First generation		3	1				N/A *		
Second generation				3	2				
Homeland Italian (n=16)									
		3	1	2	0	2	3	3	2
Heritage Faetar (n=13)									
First generation		4	4				N/A		
Second generation				2	2	1			
Homeland Faetar (n=21)									
		1	1	3	4	5	5	1	1
Total (n=61)									
		11	7	10	8	8	8	4	3

TABLE 1: Speaker distribution (age at time of recording).

The only additional criteria, besides fitting in one of these cross-sections, is that the speaker self-identify as fluent enough to participate in a conversation for an hour in their heritage language. Student interviewers, themselves native speakers of the language, are trained to rely on a potential participant's self-assessment of their ability to do this as the sole criterion for inclusion/exclusion in the study.

The speech from these one-hour (on average) conversations, elicited via sociolinguistic interview methodology (detailed in Labov 1984), provides a spontaneous speech sample in which we can analyze variable patterns at phonetic, phonological, morphological and syntactic levels. After eliciting up to an hour of conversational speech, interviewers verbally administered an ethnic orientation questionnaire and guided participants through a picture description task. The questionnaire was not administered to the Faetar speakers as we did not see how to make it comparable to the situation of the other heritage languages, where speakers are likely bilingual rather than trilingual. Additionally, it hadn't yet been implemented when the Homeland Faetar fieldwork was conducted.

Fluent speakers of each language (university students in the case of Italian, the researcher assisted by several native speakers and two linguistics students in the case of Faetar) then fully transcribed the participants' speech in the sociolinguistic interview, producing a time-aligned transcript using the ELAN annotation program (Wittenburg et al. 2006). In the case of Faetar, the speech was also translated to English. This has two benefits: it allows researchers who are not fluent in Faetar to work with the data and it facilitates text searches in a language which lacks both a standardized orthography and a complete description of phonemic vs. phonetic alternations. The documentation of spontaneous speech provides the basis for many types of analysis as well as serving as a record of the state of the language at this time.

In preparation for analysis, up to 100 consecutive examples of finite verb clauses were marked in the transcript file of each speaker. Some Faetar conversations are not

* As first generation is defined as having grown up (until at least age 18) in the home country and then having spent 20+ years in Toronto, there are no first generation speakers younger than 39.

long enough to provide this many samples.¹⁴ Each of these examples was then coded for the factors predicted to condition the variation. This work was conducted by students who are fluent, but not necessarily native, speakers of Italian. For Faetar, researchers, including undergraduate and graduate students who have strong working knowledge of related languages, worked in tandem with native speakers to code the data. This provides opportunities for all parties to learn much about the structure of the language.

Factors coded for this analysis of variable null subjects were selected based on relevant previous findings regarding null subject variation in Spanish (for an overview, see Nagy 2015). Coding these factors serves two purposes. First, it allows us to determine whether the same types of internal effects appear in these smaller languages as have been established in larger languages (though discussion of this is beyond the scope of this paper). Second, they allow us to establish effects of the social factors without bias due to skewed distributions of the types of sentences produced by each speaker or speaker group. The factors are:

- (12) person, number, and gender of the subject (simplified to the most distinctive binary contrast for Italian)
- (13) presence of a [+Argument] subject (noun or a demonstrative pronoun), for Faetar only
- (14) presence of a proclitic other than the subject, e.g., direct object, indirect object, or partitive; reduced to a binary variable (presence/absence) due to insufficient evidence of finer-grained distinctions
- (15) reference of the subject (same as that of the previous phrase, or not) for Italian, new vs. old information for Faetar
- (16) tense

In addition, effects of generation (since immigration), age, and sex were measured. This allows us to test whether the generalizations listed above in (1) are applicable to these endangered and/or lesser-studied languages. With a limited number of speakers, and limited understanding of the social structures of these communities, factors such as socioeconomic status or education level could not be considered. However, we do have heritage speakers with greater (second generation) and lesser (first generation) contact with English. Contrasting the patterns of each generation constitutes part of the analysis. While comparing speakers of different ages will show only whether change is taking place or not, comparison of generations should provide a fast-track view of language change if, indeed, contact with English is a cause of language change for this variable. Similarly, comparing speaker groups in Faeto may show if there are changes in progress in the homeland variety, though we have no firm way of distinguishing between Italian contact effects and internal change (though increasing similarity to patterns found in Italian would support the former).

¹⁴ The availability of a native speaker interviewer would likely alleviate this problem.

Mixed-effects logistic regression models with the linguistic and social predictors just described as fixed effects and speaker as a random effect were constructed in order to determine which of these factors significantly condition the presence vs. absence of null subjects in each of the four varieties. Once the best models of significant predictors were established (by including only factors with significant contributions to the model), comparison across varieties, and to similar published descriptions of English are made, following the comparative variationist method (cf. Tagliamonte 2006, Walker 2010). This allows us to first establish whether there are changes in progress in these communities and then to assess the likelihood that interference from English is the cause. It is essential to make comparison between the heritage and homeland varieties in order to ascertain that effects attributed to English are not found in the homeland varieties (which have very little contact with English). This stage of the project has been conducted by faculty and student teams working together, beginning with a graduate seminar in 2010. Neither data collection nor analysis of these languages would be possible without trained investigators who are fluent speakers of relevant language.

4.1.2.1 Results for Italian. We begin by comparing the two varieties of the larger language. Homeland Italian, based on the sample of 748 tokens, exhibits a null subject rate of 79%. Its input value in the logistic regression model shown in Table 2, a mean that is corrected for uneven distribution across contexts, is 0.84.¹⁵ Heritage Italian, in a model based on a sample of 1,045 tokens (also in Table 2), has a rate of 75%, but an input value of 0.89. If contact with English were influencing the rate of null subjects in this minority variety, we would expect it to be much lower in the heritage variety, but it is higher.

Factors	Homeland Italian			Heritage Italian		
	Factor weight	% Ø	N	Factor weight	% Ø	N
Reference of the subject						
Same	0.62	83%	402	0.63	84%	519
Switch	0.38	73%	346	0.37	67%	526
	0.24			0.30		
Proclitic						
Present	[0.48]	76%	180	0.65	86%	43
None	[0.52]	79%	568	0.35	75%	1002
				0.30		
Grammatical person (number)						
PL	0.65	89%	265	0.63	85%	350
SG	0.35	73%	483	0.37	70%	695
	0.30			0.26		
Tense						

¹⁵ Both the overall input value and the factor weights reported for each context hypothesized to influence the rate of pro-drop range from 0 to 1. In the following tables, factor weight values closer to 1 indicate a greater likelihood of null subjects than in contexts with values closer to 0.

PST PRF	[0.51]	79%	193	0.60	82%	261
N-PST	[0.57]	80%	432	0.46	73%	547
N-PRF	[0.42]	73%	123	0.43	72%	237
<i>0.17</i>						
Sex						
Female	[0.51]	76%	254	[0.41]	69%	429
Male	[0.49]	80%	494	[0.59]	80%	616
Age Group						
70-79				[0.56]	82%	147
60-69	[0.40]	75%	193	[0.45]	77%	230
50-59	[0.71]	88%	50	[0.63]	81%	369
40-49	[0.57]	83%	46	[0.36]	63%	299
30-39	[0.49]	79%	94			
20-29	[0.32]	73%	159			
10-19	[0.51]	83%	206			

TABLE 2: Two logistic regression models showing null subject variability in Heritage (input = 0.89, n = 1,045) and Homeland (input = 0.84, n= 748) Italian. Application value = null subject. Factors are listed by decreasing size of effect (range, shown in italics) in the heritage data. Factor weights for non-significant predictors are in square brackets.

We also find very little evidence of direct English interference when we consider the factors that influence the variability: reference of the subject (same as in previous phrase, or not) and number of the subject (singular or plural) have parallel effects in both varieties, as seen in Table 2. Two additional factors have significant conditioning effects in Heritage but not Homeland Italian: the presence of a proclitic and tense of the main clause verb. English lacks proclitic objects and so this effect cannot be directly attributed to contact with English.

It would be interesting to know if there is a sex or age difference to help identify how the proclitic effect originated, but with only 43 direct object examples in the corpus, such confirmation is unlikely. There is no significant difference in either the rate of null subjects or the rate of proclitic use between older and younger speakers. There is a suggestive (but non-significant) trend of a bigger effect of proclitic presence for female speakers than for male speakers. The conditioning by these social factors is illustrated in Figure 2, comparing rates of null subjects for different contexts. While the four age groups do not differ in any marked way in the size of gap between null subject rates for clauses with proclitics vs. those without, we do see that there is more of an effect (a bigger gap between the null subject rates) for females than males. This must be interpreted cautiously as the female speakers only produced a total of 15 sentences with proclitics in the sample. We cannot conclusively isolate the group of speakers who initiated this innovation, due to the rarity of proclitics.

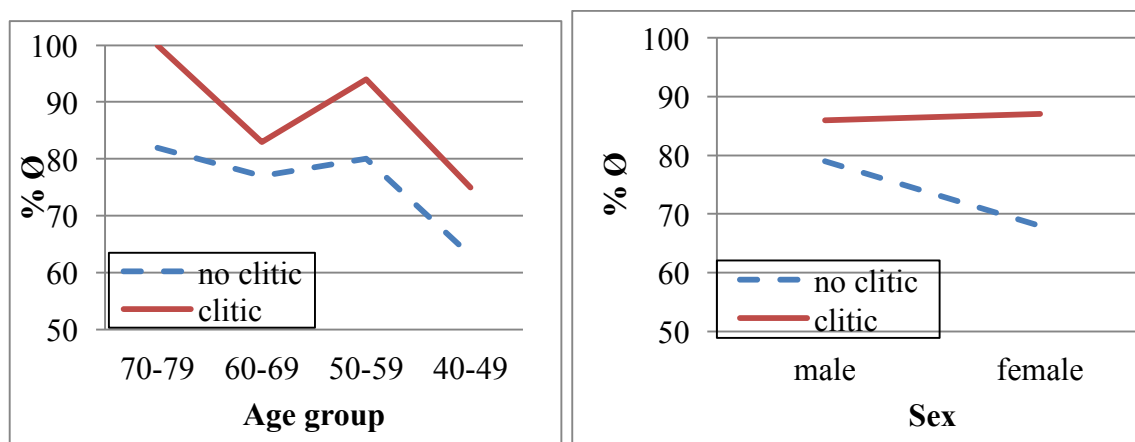


FIGURE 1: Effect of proclitic presence on null subject rate in Heritage Italian.

The other factor that behaves differently in the two Italian dialects is Tense. Heritage Italian shows a preference for more null subjects in perfect than present tense clauses, as does Toronto English. Similarly, Heritage Italian has similar null subject rates for imperfect and present tense verbs and English shows no difference in null subject rates between these (Nagy et al. 2011). Some effect from English is thus plausible for this one factor.

Neither Age nor Sex is significant in either Heritage or Homeland Italian null subject variation. In a separate model for the heritage speakers, Generation (since immigration), instead of Age, was tested but also not found significant. These are the clearest signs that no change is in progress for (pro-drop) in Heritage Italian. If there is no change, there can be no evidence of contact from English.

We can see this another way as well. The analyses that produced the models in Table 2 also included Speaker as a random effect. This provides us with the mean rate for each speaker, showing the range of inter-speaker variation, as well as preventing any speakers with extreme rates from over-influencing the probability (factor weight) of the groups they are in. The stability of this variable between younger and older speakers can be seen in Figure 2, in which the proportion of null subjects are plotted against age by individual speaker. We also see the similarity in rate and range between homeland and heritage speakers, and between males and females.¹⁶ While there are four speakers with noticeably lower rates of pro-drop than the others, they do not have anything obviously in common: they are not all the same generation, nor the same age, nor the same sex, nor interviewed by the same interviewer, nor do they have similar response patterns in the ethnic orientation questionnaire.

¹⁶ A plot with factor weights instead of percentages would show the same results.

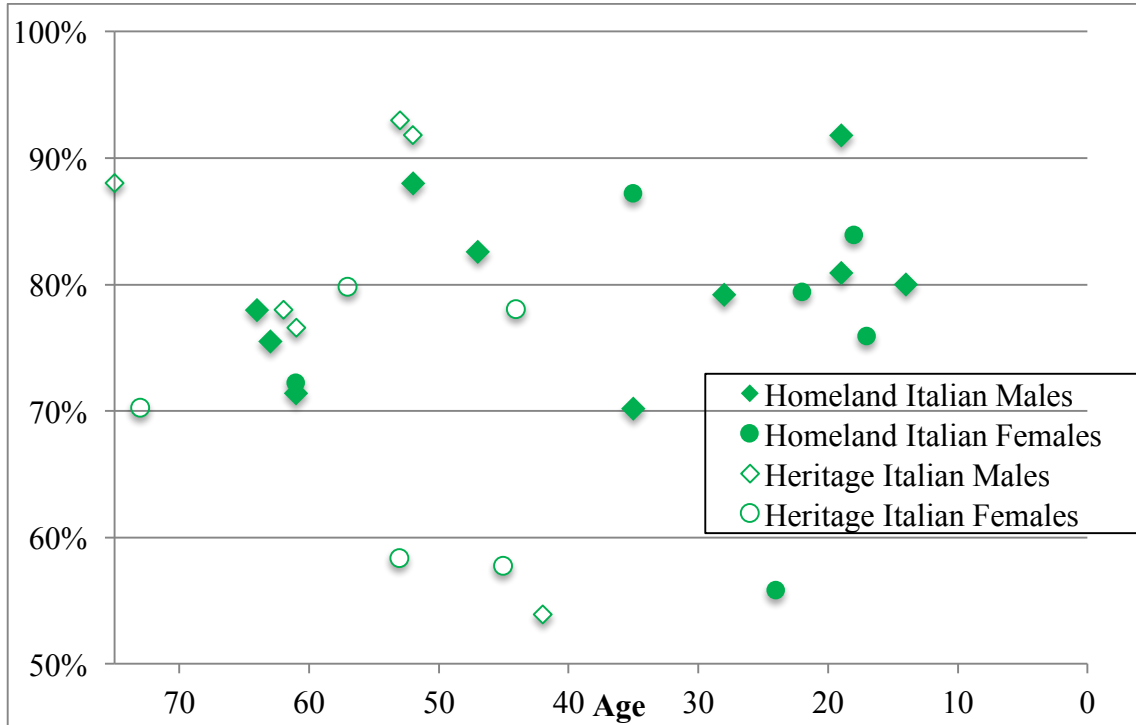


FIGURE 2: Null subject rates by individual (Homeland and Heritage Italian)

4.1.2.2 Results for Faetar. We turn next to Faetar, again comparing the homeland and heritage varieties in terms of both rates and conditioning effects. First, the overall rate of null subject differs: 55% (input value .37) in Homeland Faetar but 39% (input value .40) in Heritage Faetar. However, there is again strong similarity between the Homeland and Heritage varieties in terms of conditioning effects. The same conditioning factors exhibit the strongest effects in both varieties: Grammatical Person, [+Argument] Subject, and Proclitic (see Table 3). There are slight differences in the ordering of the factors within Grammatical Person only where data is scarce. Tense has the weakest grammatical effect in the Homeland data and is not significant in the model of the Heritage Faetar data. The effect of Information Status is also consistent, and weak, in both varieties.

Factor group	Homeland Faetar			Heritage Faetar		
	Factor weight	% Ø	N	Factor weight	% Ø	N
Grammatical person						
Generic	0.97	90	561	0.84	66	55
3SG	0.72	44	780	3M 0.76	57	462
3PL	0.57	25	203	3F 0.67	44	111
1PL	0.11	6	32	0.40	19	47
2SG	0.49	17	12	0.19	7	92
1SG	0.06	3	62	0.06	6	192
	<i>91</i>			<i>69</i>		
[+Argument] Subject						
Demonstrative	0.76	75	231	0.67	78	50
			116			
None	0.42	57	2	0.35	34	801
Noun	0.30	28	257	0.48	61	108
	<i>46</i>			<i>32</i>		
Proclitic						
PRS	0.73	82	147	0.58	62	55
			150			
None	0.27	52	3	0.42	38	904
	<i>47</i>			<i>17</i>		
Information Status						
			126			
New	0.59	61	1	0.55	42	649
Old	0.41	35	373	0.45	30	294
	<i>19</i>			<i>10</i>		
Age Group						
Oldest	0.61	60	434	0.61	49	198
Middle	0.51	60	681	0.45	37	494
Youngest	0.38	44	535	0.44	35	267
	<i>23</i>			<i>16</i>		
Tense						
			156			
N-PST	0.63	56	0	[0.47]	37	480
PST	0.37	37	90	[0.53]	41	479
	<i>26</i>					

TABLE 3: Null subject variability in Homeland (n=1,580; Input =0.37) and Heritage Faetar (n=942; Input = 0.40) Application value = null subject. Factors are listed by decreasing size of effect (range, shown in italics) in the heritage data. Factor weights for non-significant predictors are in square brackets.

In Faetar, new information is more likely to be presented with a null subject than old/given, an unusual effect cross-linguistically. At first glance, this seemed easily accounted for: as noted above, Faetar subject pronouns and clitics may appear with Noun

subjects (rather than being in complementary distribution to nominal subjects, as in Italian). Thus, clauses coded as having a null subject, but which have a nominal subject, could introduce new material. However, such an explanation is not viable: if we consider only sentences with no [+Argument] subject, we still see higher rates of null subjects when introducing new than old information (see Table 4). It is interesting that the difference in null subject rates for new vs. old information is considerably smaller for the heritage than the homeland variety.

	Homeland Faetar		Heritage Faetar	
	% Ø-subject	N	% Ø-subject	N
New information	67%	824	36%	510
Old information	32%	328	27%	264

TABLE 4: Faetar null subjects for new information (contexts with no [+Argument] subject only)

Age has a consistent effect in the two communities while Sex is significant in neither. Both the heritage and homeland Faetar varieties show a decrease in the use of null pronouns among younger speakers, but no effect of sex. see Figure 3.¹⁷ The Heritage Faetar speakers were all shifted right on the x-axis by 20 years in order to account for the gap between recording heritage and homeland speakers. Both by birth year and in terms of range of pro-drop rates, they overlap almost exactly with the Homeland Faetar group. Again, Generation had no significant effect in the heritage data when tested in a similar model (where Age was excluded).

¹⁷ Figure 3 presents the raw percentage of tokens with null subjects. This correlation to age is not a strong one ($R^2=0.35$ for Heritage Faetar and $R^2=0.30$ for Homeland Faetar). However, age is a significant factor in the model in Table 3 because the model extrapolates the effect of age from a complex combination of factors influencing variant selection. While the model reflects multiple influencing factors simultaneously, the raw percentages do not.

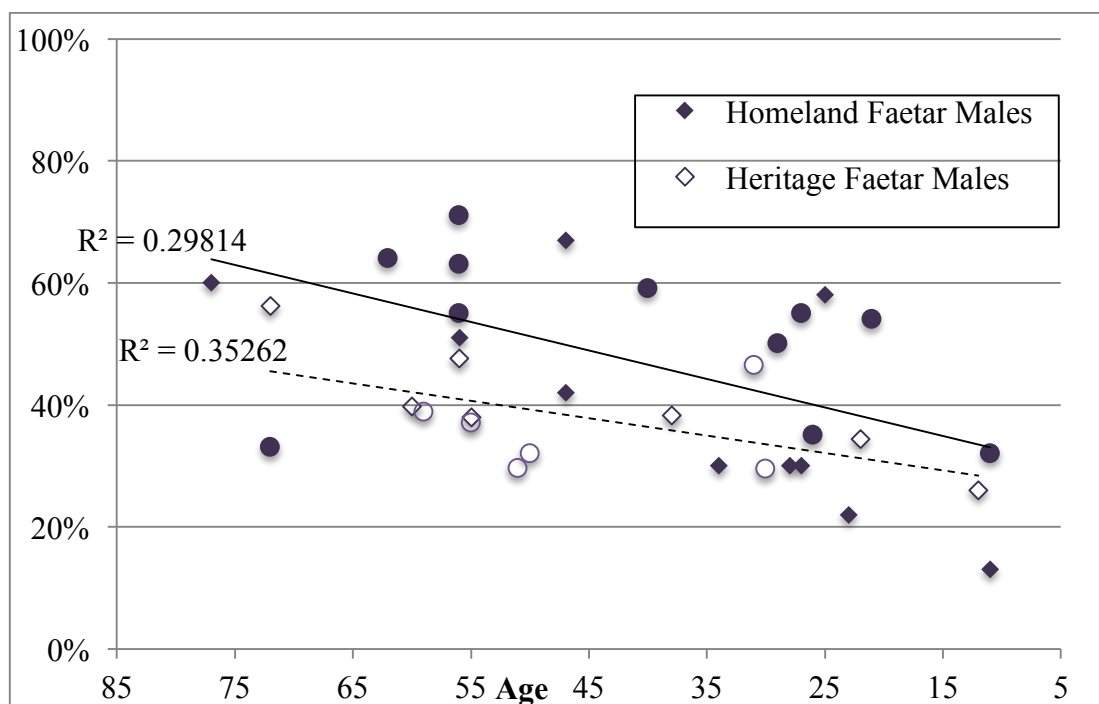


FIGURE 3: Null subject rates by individual (Homeland and Heritage Faetar). Note that Heritage speakers have been shifted to the right by 20 years to account for the 20-year gap between recording heritage and homeland speakers.

In terms of contributions by and benefits to students, it is worth mentioning that much of the statistical analysis for both Italian and Faetar data has been carried out by students with limited or no fluency in the language, after careful checking with native speaker assistants about the meaning of the data they work with. This gives students an opportunity to expand their methodological skills for documentation work and increase their knowledge in the field of sociolinguistics by presenting new analyses of previously undocumented varieties. Students presented pilot studies for most of the results presented here at conferences and have co-Nagyed some publications, extending their skill acquisition to presentation and dissemination.

5. DISCUSSION: COMPARING MORE AND LESS ENDANGERED LANGUAGES. This section discusses the patterns of variation for variable null subjects in these four varieties as examples of the contributions to our understanding of sociolinguistic variation, and the factors that contribute to it, that may be achieved by collaborative work between native speakers of minority varieties and trained sociolinguists.

We first discuss the data for the two varieties of Italian. There are two surprising results, given the different contexts in which the varieties have developed. The first is how similar the heritage and homeland varieties are in terms of rate of null subjects, stability across sex and age groups, and primary conditioning factor effects. Reference of the Subject and Grammatical person (Number) have virtually identical effects in the two data sets, despite one variety being produced in an English-dominant community in a city

where ~3% of the population claims Italian as a mother tongue and the other in a community where everyone speaks Italian and few speak English.

The other surprise lies in the differences that do exist between the two varieties: the heritage variety has a more complex grammar, exhibiting conditioning effects of Proclitics and Tense. While we might interpret the tense effect as due to influence from English, as noted in the previous section, the innovation of a proclitic effect must be *sui generis*. This effect is noteworthy as it contradicts expectations of grammar simplification in minority language communities.

Finally, note that the rate of null subjects, overall, is *slightly* lower in the Heritage Italian data than in the Homeland Italian data. While, as an isolated fact, this might suggest English influence, we will return to this issue after considering the Faetar data.

The pattern of younger Faetar speakers using fewer Ø-subjects suggests a shift away from an Italian-like grammar, perhaps a response to its high degree of contact with Italian both in Apulia and in Toronto. The motivation of marking a distinctive identity is plausible in both places, where Faetar speakers may feel overshadowed by the large Italian-speaking community. The lack of a generational difference on the rate in Heritage Faetar indicates that it is unwise to attribute the change to contact with English: the second generation, born and raised in Toronto, have more contact with English, but do not have a significantly different Ø-subject rate from Generation 1 speakers. Rather, the Heritage speakers are clearly continuing a trend established in the homeland, where English is an unlikely scapegoat.

Having seen that this is the case for Faetar, it is harder to make the case that the (very slightly) lower rate of null subject for Heritage Italian than Homeland Italian must be attributed to English. Again, there is no effect of generation. We would expect Generation 2 speakers to have significantly lower rates of null subjects than Generation 1 if language use were to be called to account for the change.

Careful examination of trends in contexts of greater and less contact with other languages, and concomitant greater and lesser minority status, can shed new light on patterns that might otherwise be attributed to the influence of a globally dominant language. In this case, the endangered language's behavior has helped interpret the larger language's trends, suggesting that there must be an alternative explanation to influence from English to account for the differences seen between Homeland and Heritage varieties of Italian. Rather, we have provided evidence for (internal) change in four varieties that have been subject to very little prior variationist analysis. Sex was not found to have a significant effect in any analysis of this pattern exhibiting ongoing changes. This runs counter to expectation, given the generalizations about sex effects in (1).

6. CONCLUSION. We have seen that in both Italy and Toronto, although the tiny Faetar communities are expected to be under intense contact pressure from English and/or Italian; however, this does not appear to be the case. Rather, Faetar is resisting the pulls in all directions from its Francoprovençal roots: from Italian in its homeland and from its current contact with English in Toronto. Heritage Faetar remains very similar to Homeland Faetar: both show a slight trend toward more overt subjects, and similar constraint effects, suggesting a pull away from Italian. However, this cannot be considered strong evidence of a pull toward English, because it is also occurring on a

mountainside in rural Puglia (where English is still rarely used, and was virtually absent in the 1990's when these speakers were recorded).

Heritage Italian retains a very similar Ø-subject rate to its homeland counterpart, but is more complex than the Homeland variety in that it has more significant effects constraining its variability. Again, these cannot be easily attributed to English.

This set of findings provides a competing perspective to the expectation that simplification and adoption of features of the majority language are the necessary outcomes for minority and endangered languages, a view exemplified, for example, by this passage from Fernández-Ordóñez (2012:73-4):

The loss of previously existing distinctions seems to occur more easily in social situations where speakers of different languages or dialects colonize new territories, bringing their varieties into contact...Simply put, dialect contact usually implies altered replication, since the structural constraints are not wholly acquired in contact between adults.

Through a collaborative project involving contributions from native speakers, students (sometimes one and the same) and professional scholars, we see, instead, that endangered, small, minority languages may exhibit internal changes toward greater complexity and the same sort of orderly heterogeneity frequently observed in large languages. This is one example of the type of new insight offered by expansion of variationist methods to a wider set of languages and community settings.

REFERENCES

- Albirini, Abdulkafi & Brahim Chakrani. 2016. Switching codes and registers: An analysis of heritage Arabic speakers' sociolinguistic competence. *International Journal of Bilingualism*. doi: 10.1177/1367006915626587.
- Bailey, Guy, Tom Wikle, Jan Tillery & Lori Sand. 1991. The Apparent time construct. *Language Variation and Change* 3 (3). 241-65.
- Bayley, Robert. 2013. Variationist sociolinguistics. In Robert Bayley, Richard Cameron and Ceil Lucas, eds. 2013. *Oxford Handbook of Sociolinguistics*. 11-30. NY: Oxford University Press.
- Bayley, Robert, Richard Cameron & Ceil Lucas (eds.). 2013. *Oxford Handbook of Sociolinguistics*. NY: Oxford University Press.
- Bell, Allan. 1984. Language style as audience design. *Language in Society* 13.145-204.
- Blainey, Darcie. 2013. *First to come, last to go: Phonological change and resilience in Louisiana regional French*. New Orleans, LA: Tulane University dissertation.
- Cedergren, Henrietta J. 1973. *The interplay of social and linguistic factors in Panama*. Ithaca, NY: Cornell University dissertation.
- Cedergren, Henrietta J. 1988. The spread of language change: Verifying inferences of linguistic diffusion. In Peter Lowenberg (ed.), *Language spread and language policy: Issues, implications, and case studies*, 45-60. Washington, DC: Georgetown University Press.
- Childs, Tucker, Jeff Good & Alice Mitchell. 2014. Beyond the ancestral code: towards a model for sociolinguistic language documentation. *Language Documentation & Conservation*. 8. 168-191. <http://hdl.handle.net/10125/24601>
- Chirkova, Katia, James Stanford & Dehe Wang. 2016. Documenting consonantal variation in Ganluo Ersu. Paper presentation in the "Documenting variation in endangered languages" symposium, Linguistic Society of America Annual Meeting, 7-10 January, Washington DC.
- Cieri, Christopher. 1985. *Italian lexical items in the English speech of Italo-Americans*. Philadelphia, PA: University of Pennsylvania MA thesis.
- City Population. 2016. <http://www.citypopulation.de/>. (13 May, 2016.)
- Danesi, Marcel. 1983-84. Canadian Italian as a Marker of Ethnicity. *NEMLA Italian Studies* 7-8:99-105.
- Danesi, Marcel. 1984. Canadian Italian: A Case in point of how language adapts to environment. *Polyphony* 7. 110-113. Multicultural History Society of Ontario.
- D'Arcy, Alexandra. 2013. Variation and change. In Robert Bayley, Richard Cameron & Ceil Lucas (eds.). *The Oxford Handbook of Sociolinguistics*. 484-502. Oxford University Press.
- Del Torto, Lisa. 2008. Once a broker, always a broker: Non-professional interpreting as identity accomplishment in multigenerational Italian-English bilingual family interaction. *Multilingua* 27. 77-97.
- Del Torto, Lisa. 2010. 'It's so cute how they talk': Stylized Italian English as sociolinguistic maintenance. *English Today* 26. 55-62.
- Dorian, Nancy. 1980. Language shift in community and individual: The phenomenon of the laggard semi-speaker. *International Journal of the Sociology of Language* 25. 85-94.

- Drager, Katie, Bethany Kaleialohapau‘ole Chun Comstock, & Hina Puamohala Kneubuhl. 2016. Ka puana pa‘a ‘ole: glottal stop variation in Hawaiian. Paper presented in the “Documenting variation in endangered languages” symposium, Linguistic Society of America Annual Meeting, 7-10 January, Washington DC.
- Dubois, Sylvie & Sibylle Noetzel. 2005. Intergenerational pattern of interference and internally-motivated changes in Cajun French. *Bilingualism: Language and Cognition* 8(2). 131–143.
- Fernández-Ordóñez, Inés. 2012. Dialect areas and linguistic change. In Gunther De Vogelaer and Guido Seiler (eds.), *The dialect laboratory: Dialects as a testing ground for theories of language change*. 73-106. Amsterdam: John Benjamins.
- Foley, William. 2005. Personhood and linguistic identity, purism and variation. In Peter K. Austin (ed.), *Language Documentation and Description*, vol. 3. 157-180. London: SOAS.
- Fortier, Anne-Marie. 1991. *Langue et rapports sociaux: analyse des langues d'usage chez des italiens de deuxième génération*. Québec: Centre international de recherche en aménagement linguistique.
- Garrett, Peter. 2010. *Attitudes to Language*. Cambridge: Cambridge University Press.
- Giles, Howard, Richard Y. Bourhis & Donald M. Taylor. 1977. Towards a theory of language in ethnic group relations. In Howard Giles (ed.), *Language, ethnicity and intergroup relations*. 307-348. London: Academic.
- Grimm, D. Rick. 2015. *Grammatical variation and change in spoken Ontario French: The subjunctive mood and the expression of future temporal reference*. Toronto, ON: York University dissertation.
- Guy, Gregory. 2015. Are there universal constraints on sociolinguistic variation? Paper presented at the Towards Cross-Cultural Sociolinguistics workshop, University of Cologne.
- Harrison, K. David. 2005. Ethnographically informed language documentation. In Peter K. Austin (ed.), *Language Documentation and Description*, vol. 3. 22-41. London: SOAS.
- Hilton, Nanna Haug & Charlotte Gooskens. 2013. Language policies and attitudes towards Frisian in the Netherlands. In Charlotte Gooskens & Renée van Bezooijen. *Phonetics in Europe: Perceptions and production*. 139-157. Frankfurt: Peter Lang.
- Holmes, Janet & Miriam Meyerhoff. 1999. The community of practice: Theories and methodologies in the new language and gender research. *Language in Society* 28:173-183.
- ISTAT [Istituto Nazionale di Statistica]. 2007. La lingua italiana, i dialetti e le lingue straniere (2006). http://www.istat.it/salastampa/comunicati/non_calendario/20070420_00/ (accessed 1 May, 2010).
- Kang, M. Agnes. 2015. Social aspects of Korean as a heritage language. In Lucien Brown and Jaehoon Yeon, eds. *The Handbook of Korean Linguistics* 405-418. John Wiley & Sons, Inc, Hoboken, NJ.
- King, Ruth. 1989. On the social meaning of linguistic variability in language death situations: Variation in Newfoundland French. In Nancy Dorian (ed.), *Investigating obsolescence: Studies in language contraction and death*. 139–148.

- Cambridge University Press.
- Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 1984. Field methods of the project on linguistic change and variation. In John Baugh & Joel Sherzer (eds.). *Language in Use*. 28-53. Englewood Cliffs: Prentice Hall.
- Labov, William. 1990. The interaction of sex and social class in the course of linguistic change. *Language Variation and Change* 2:205-254.
- Labov, William. 2001. *Principles of Linguistic Change. Vol.2. Social factors*. Malden, MA: Blackwell.
- Labov, William. 2015. The discovery of the unexpected. *Asia-Pacific Language Variation* 1. 7-22.
- Lealess, Allison & Chelsea T. Smith. 2011. Assessing contact-induced language change: The use of subject relative markers in Quebec English. *Cahiers Linguistiques d'Ottawa* 36.
<http://aix1.uottawa.ca/%7Eclo/36/Lealess.Smith.2011.pdf> (accessed 10 October 2016).
- Le Page, Robert & Tabouret-Keller, Andrée. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
- Leroux, Martine and Lidia-Gabriela Jarmasz. 2006. A Study about nothing: Null Subjects as a diagnostic of convergence between English and French. *Penn Working Papers in Linguistics* 12(2). 1-14.
- McMullen, Kathryn. 2011. Postsecondary education participation among underrepresented and minority groups. Statistics Canada.
<http://www.statcan.gc.ca/pub/81-004-x/2011004/article/11595-eng.htm#a>. (7 April, 2016.)
- Meyerhoff, Miriam. *forthcoming*. Writing a linguistic symphony: Analyzing variation while doing language documentation. *Canadian Journal of Linguistics*.
- Meyerhoff, Miriam. 2016. Borrowing from Bislama into Nkep (East Santo, Vanuatu): Quantitative and qualitative perspectives. *Language & Linguistics in Melanesia* 34(1). 77-94.
- Meyerhoff, Miriam & Naomi Nagy (eds.). 2008. *Social Lives in Language -- Sociolinguistics and multilingual speech communities*. John Benjamins: Amsterdam.
- Milroy, James. 2001. Language ideologies and the consequences of standardization. *Journal of Sociolinguistics* (5). 530-55.
- Milroy, Lesley & James Milroy. 1992. Social network and social class: Toward an integrated sociolinguistic model. *Language in Society* 21:1-26.
- Montrul, Silvina. 2009. Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism* 13.2:239-269.
- Nagy, Naomi. 2000. *Faetar*. Munich: Lincom Europa.
- Nagy, Naomi. 2009. The challenges of less commonly studied languages: Writing a sociogrammar of Faetar. In James Stanford & Dennis Preston. (eds.). *Variation in Indigenous Minority Languages*. 397-417. Philadelphia: John Benjamins. *Impact series*, vol. 25.
- Nagy, Naomi. 2011. A multilingual corpus to explore geographic variation. *Rassegna*

- Italiana di Linguistica Applicata* 43(1-2). 65-84.
- Nagy, Naomi. 2015. Evidence for change vs. incomplete acquisition in Toronto Heritage Languages. In P. Guijarro-Fuentes & K. Schmitz, eds. *Fundamentally (in)complete grammars? Emergence, acquisition and diffusion of new varieties*. *Lingua* 164B:309-327.
- Nagy, Naomi. 2016. Heritage languages as new dialects. *The Future of Dialects: Select papers from Methods in Dialectology XV*. Marie-Hélène Côté, Remco Knooihuizen & John Nerbonne (eds.). 15-34. Berlin: Language Science.
- Nagy, Naomi. 2017. Heritage language speakers in the university classroom, doing research. In P. Trifonas & T. Aravossitas, eds. *International Handbook on Research and Practice in Heritage Language Education*. Springer. doi: 10.1007/978-3-319-38893-9_41-1.
- Nagy, Naomi, Joanna Chocieł & Michol Hoffman. 2014. Analyzing Ethnic Orientation in the quantitative sociolinguistic paradigm. In Lauren Hall-Lew & Malcah Yaeger-Dror (eds.). Special issue of *Language and Communication: New perspectives on the concept of ethnolect*. 35. 9-26.
- Nagy, Naomi & Alexei Kochetov. 2013. VOT across the generations: A cross-linguistic study of contact-induced change. In Peter Siemund, Ingrid Gogolin, Monika Schulz & Julia Davydova (eds.). *Multilingualism and language contact in urban areas: Acquisition - Development - Teaching - Communication*. 19-38. John Benjamins.
- Nagy, Naomi, Nina Aghdasi, Derek Denis & Alexandra Motut. 2011. Pro-drop in Heritage Languages: A cross-linguistic study of contact-induced change. *Penn Working Papers in Linguistics* 17(2).
- Nagy, Naomi & Miriam Meyerhoff. 2008. The social lives of linguistics. In Miriam Meyerhoff & Naomi Nagy (eds.). *Social Lives in Language -- Sociolinguistics and multilingual speech communities*. 1-17. Amsterdam: John Benjamins.
- Nagy, Naomi, Michael Iannozzi & David Heap. *forthcoming*. Faetar null subjects: A variationist study of a heritage language in contact. *International Journal of the Sociology of Language* 249.
- Nagy, Naomi & Yoonjung Kang. 2016. VOT Merger in Heritage Korean in Toronto. *Language Variation and Change* 28.2:249-272.
- Nichols, Johanna. 2008. Universals and diachrony: Some observations. In Jeff Good (ed.), *Linguistic Universals and Language Change*. 287-293. Oxford: Oxford University Press.
- Ó Riagáin, Pádraig. 2008. Language attitudes and minority languages. In Nancy Hornberger, ed. *Encyclopedia of language and education*. 2077-2089. Springer US.
- Palumbo, Maria. 2014. *Cultura, lingua, identità: Uno studio sul parlato di emigrati calabresi*. Roma: Aracne.
- Polinsky, Maria. 2011. Reanalysis in adult heritage language: a case for attrition. *Studies in Second Language Acquisition* 33.2:305-328.
- Poplack, Shana & Stephen Levey. 2010. Contact-induced grammatical change: A cautionary tale. In Peter Auer & Jürgen Schmidt (eds.). *Language and Space: An International Handbook of Linguistic Variation*. Mouton de Gruyter.
- Poplack, Shana & Marjory Meechan. 1998. How languages fit together in codemixing.

- International Journal of Bilingualism* 2(2). 127-38.
- Poplack, Shana & Danielle Turpin. 1999. Does the FUTUR have a future in (Canadian) French?. *Probus* 11. 133-164.
- Poplack, Shana, James A. Walker, & Rebecca Malcolmson. 2006. An English "like no other"? Language contact and change in Quebec. *Canadian Journal of Linguistics* 51(2/3). 183-213.
- Ravindranath, Maya. 2015. Sociolinguistic variation and language contact. *Language and Linguistic Compass* 9(6). 243-55.
- Rice, Keren. 2014. On beginning the study of the tone system of a Dene (Athabaskan) language: Looking back. *Language Documentation & Conservation* 8. 690-706. <http://hdl.handle.net/10125/24621>
- Rottet, Kevin. 1995. *Language shift and language death in the Cajun French-speaking communities of Terrebonne and Lafourche Parishes, Louisiana*. Bloomington, IN: Indiana University dissertation.
- Sankoff, Gillian. 2013. Longitudinal studies. In Robert Bayley, Richard Cameron and Ceil Lucas, eds. *Oxford Handbook of Sociolinguistics*. 261-279. NY: Oxford University Press.
- Stanford, James. 2013. How to uncover social variables: A focus on clans. In *Data collection in sociolinguistics: Methods and applications*, Christine Mallinson, Becky Childs & Gerard Van Herk (eds.). 25-28. Routledge Press.
- Stanford, James & Dennis Preston (eds.). 2007. *Variation in Indigenous Minority Languages*. Philadelphia: John Benjamins. *Impact* series, vol. 25.
- Statistics Canada. 2011a. Population by mother tongue. <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo12c-eng.htm>. (15 May, 2016.)
- Statistics Canada. 2011b. Population by selected ethnic origins. <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo27g-eng.htm>. (15 May, 2016.)
- Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Tagliamonte, Sali & Alexandra D'Arcy. 2007. Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change* 19(2). 199-217.
- Thepboriruk, Kanjana. 2015. *Thai in Diaspora: Language and identity in Los Angeles, California*. Mānoa, HI: University of Hawai'i dissertation.
- Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh University Press.
- Thomason, Sarah G. & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley CA: University of California.
- Torres Cacoullos, Rena & Catherine Travis. 2015. Gauging convergence on the ground: Code-switching in the community. *International Journal of Bilingualism* 19(4). 365-386.
- Walker, James A. 2010. *Variation in Linguistic Systems*. New York and London: Routledge.
- Walker, James A. & Miriam Meyerhoff. 2013. Studies of the community and the individual. In Robert Bayley, Richard Cameron and Ceil Lucas, eds. *Oxford Handbook of Sociolinguistics*. 175-194. NY: Oxford University Press.

- Waltermire, Mark. 2010. Variants of intervocalic /d/ as markers of sociolinguistic identity among Spanish-Portuguese bilinguals. *Spanish in Context* 7(2). 279–304.
- Wei, Li. 2013. Conceptual and methodological issues in bilingualism and multilingualism research. In Tej Bhatia & William Ritchie (eds.). *The Handbook of Bilingualism and Multilingualism*. 26-51. Malden, MA: Wiley-Blackwell.
- Wittenburg, P., H. Brugman, A. Russel, A. Klassmann, A., & H. Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation*, 1556.
- Woolard, Kathryn. 1989. Language convergence and language death as social processes. In Nancy Dorian (ed.). *Investigating obsolescence: Studies in language contraction and death*. 355-367. New York, NY: Cambridge University Press.

naomi.nagy@utoronto.ca